Xcel Energy℠

414 Nicollet Mall
Minneapolis, Minnesota 55401-1993

September 24, 2007

Mr. Scott Hansen
Remedial Response Branch, Region 5
U. S. EPA (SR-6J)
77 West Jackson Blvd.
Chicago, IL 60604

RE: Responses to USEPA's Required Changes to the Baseline Ecological Risk Assessment (BERA) - Ashland/NSP Lakefront Superfund Site

Dear Mr. Hansen:

USEPA required that several changes be made to the conclusions of the site-specific benthic community study. These changes materially changed the conclusions of BERA. USEPA supported this directive with three categories of comments that were provided in USEPA's Technical Memorandum discussing the Sediment Preliminary Remediation Goal (PRG) for the Ashland/NSP Lakefront site submitted under USEPA letter dated April 25, 2007. Comments in USEPA's Technical Memorandum generally fell into the following three categories:

1) Since there was substantial variability in the benthic community structure at both Site and reference stations, the results of statistical analyses of the benthic community data could not be relied upon.
2) The sand reference stations used for the benthic community study were not appropriate because:
    a. There was significant mortality associated with the sediments from these stations in the sediment toxicity tests; and
    b. Sediments from SQT10 and SQT12 exhibited, "a strong odor of decaying organic matter" and "elevated levels of ammonia" while in the sediment toxicity laboratory.
3) NSPW erred by placing a high weight of evidence on the results of the benthic community study because according to the EPA's Guidance Manual to Support the Assessment of Contaminated Sediments in Freshwater Ecosystems – Volume III – Interpretation of the Results of Sediment Quality Investigations (EPA-905-BO2-001C, December 2002), "the information on benthic community structure can not be used alone to evaluate the cause of any impacts observed." [and] "The "high weight of evidence" that NSPW attempts to place on the benthic macroinvertebrate community study is not supported by USEPA guidance".

Based upon these comments, USEPA concluded that the results of the benthic community study had no value and totally ignored this line of evidence in developing a PRG for Site sediments. Only the sediment chemistry and toxicity data were used as the basis of the sediment PRG.

USEPA explicitly stated that,

> *"With the variability and uncertainty outlined above* [in the benthic community study], *the statistical analysis of the community data is questionable; as such, it* [the benthic community study] *was not used to derive the proposed PRG* [emphasis added],[1]

USEPA's conclusion to totally ignore the benthic community study as a line of evidence led to USEPA requiring the following additions to and deletions from the BERA:

> Page ES-11 (refers to page number in the BERA):
> (Added) *Effects observed from field surveys of the existing benthic community indicated effects that were less dramatic than those demonstrated in the laboratory toxicity studies, but interpretation of the field survey data is made very difficult by a high degree of variability and lack of comparability between reference and site stations.*
>
> Page 5-24:
> (Added) *However, there was tremendous variability and resultant uncertainty associated with both the site samples and reference samples collected in the benthic macroinvertebrate community investigation.*
>
> Page 6-14:
> (Deleted) *There is no apparent impact to benthic macroinvertebrate communities, even in Site area, based upon the absence of any significant structural difference between Site stations and reference stations.*
>
> Pages 6-26 and 6-27:
> (Added) *However, there is uncertainty associated with the reference locations that produces questionable results and yields low power, including, but not limited to:*
>
> *The reference sand sites SQT10 and SQT12 exhibited "a strong odor of decaying organic matter" and "elevated levels of ammonia [in laboratory samples]";*
>
> *The reference sand sites SQT10 and SQT12 exhibited <50% survival for Hyalella azteca 28 day sediment exposure toxicity test;*
>
> *The reference wood site SQT11 had no survival in several replicates of the Lumbriculus bioaccumulation study;*
>
> *The reference sand sites SQT13 and SQT 14 were collected in Fall 2005 versus Spring 2005, more than 3 months after the initial sample collection. Use of this data is questionable for comparison of population metrics due to expected seasonal variation in larval and emergent species; and*

1 See page 7 in EPA's "Technical Memorandum on the Derivation of Sediment Preliminary Remediation Goal (PRG) for the Ashland Lakefront Site" (April 9, 2007).

Only three site locations appear to be "sand" sites, and none of the reference sand sites appear to be appropriate. Thus, the sample size for sand sediments does not appear meet the power requirements outlined in the RI/FS workplan.

Page 7-1:
(Added) *Effects observed from field surveys of the existing benthic community indicated effects that were less dramatic than those demonstrated in the laboratory toxicity studies, but interpretation of the field survey data is made very difficult by a high degree of variability and lack of comparability between reference and site stations.*

USEPA required these changes despite the fact that rigorous statistical analyses, analyses approved by USEPA as part of the RI/FS Work Plan, concluded that there were no apparent differences in the benthic invertebrate community at Site or reference stations that could be attributed to presence of PAHs in Site sediments.

NSPW's Response to USEPA's Required Changes to the BERA

The following provides NSPW's responses to the general themes of the USEPA-required changes to the BERA:

I. Since there was substantial variability in the benthic community structure at both Site and reference stations, the results of statistical analyses of the data could not be relied upon.

Starting with the undisputed premise that there was substantial variability in the benthic community structure at both Site and reference stations, USEPA used a specious argument that, because of this variability, there was substantial uncertainty associated with the statistical analysis and as a result the conclusions of the benthic community report could not be relied upon. However, as was explained in the RI/FS Work Plan (URS 2005), in light of the inherent variability in benthic community structural characteristics care was taken in the design of the benthic community studies to get adequate replication of samples. The experimental design for sampling proposed in the RI/FS Work Plan and approved by USEPA was supported by power analysis using historical SEH benthic community data to ensure that any differences in the benthic community study that were due to the presence of PAHs could be differentiated with reasonable statistical confidence from natural variability. The experimental design for benthic sampling as well as the power analysis and proposed statistical analytical methods were reviewed and approved by USEPA.

The approach to sampling and analysis of the benthic community data was explained further in NSPW responses to the initial USEPA comments (October 27, 2006) to the Draft BERA dated May 30, 2006 about variability of benthic community structure, to wit:

> USEPA Comment 16. There is a heavy emphasis on the benthic community study as being the strongest line of evidence and not providing clear evidence of in situ effects on the benthic community. While the conceptual rationale for this is reasonable, it assumes the study has the discriminatory power to detect differences. The degree of variability observed, both within and between sampling locations, brings this very much into question. If the benthic community study has low power, then it is prone to underestimating effects and is in fact a weak line of evidence rather than a strong one.

NSPW Response

There is indeed great variation among sampling locations, and this is why we removed ("partialed" out) the variation due to sampling location which included substrate effects in our Analysis of Covariance (ANCOVAs) before looking for effects of PAH. Our null hypothesis was that including carbon normalized PAHs (NOCPAH) and total PAHs (TPAH) in the model produced no increase in R^2 (variation explained) after removing the variation explained by among station variability. Cohen (1988) posits increase in R^2 of 0.02, 0.13, and 0.26 as 'small', 'medium', and 'large' effect sizes, respectively. Using these effect sizes in Cohen's own program, SamplePower 2.0®, with 1, 55 degrees of freedom (df) our analysis reveals an a priori power to reject the null hypothesis when the alternative hypothesis of a 'medium' effect is true of 0.819. The power to detect a 'large' effect is 0.993 or almost certainty. So based upon these results if PAHs are having any effect on community structure, it is small. Power to detect a 'small' effect was 0.182,

This quite respectable power of over 0.8 to detect a 'medium' effect refutes the argument that the "...study... is ... a weak line of evidence rather than a strong one." Our study and analyses did not have low power to detect effects of PAH on benthic community structure.

To further support our contention, we performed an a posteriori power analysis on the result of our ANCOVAS. While such analyses have limited usefulness, they can shed light on the frequency with which high variability makes detection of real effects difficult, using the actual partitioning of variability into that explained by PAH concentration (after removing among station variability) and unexplained variability. Table 2 presents the result of that analysis: power to detect a 'small' effect after partialing out station variability varied from a low of 0.34 to as high as 0.94. Average power was 0.57, or better than one chance in two of rejecting a false null hypothesis. Power to detect a 'medium' effect was 1.0: certainty for all 22 biological metrics! Again this indicates that the power of these statistical analyses to detect a change due to PAH was substantial.

What effect sizes were actually observed in this study? The third and last column in Table 2 presents these for the effect of both NOCPAH and TPAH, respectively. Average increase in R^2 for including NOCPAH in the model was 0.005; that for including TPAH in the model was 0.0036. The actual effect sizes observed were in general much smaller than Cohen's standard for a 'small' effect. Again, we argue that if PAHs are affecting community structure, the effect is small. Since we had good power to detect a 'medium' effect, failure to reject the null hypothesis in most of the ANCOVAs was not due to low power.

One may also argue that analyzing 22 different metrics of community structure constitutes a form of meta-analysis. Our repeated inability to detect an effect spread over a number of different measures of community structure further suggests the effect of PAHs on benthic community structure, if any, has to be small.

Over the next six months USEPA failed to rebut, discuss, or, in fact, even acknowledge NSPW's response in any subsequent meeting or discussion despite several NSPW requests to discuss the benthic community study. Meanwhile, during this same period, a considerable technical dialogue between NSPW's consultants and USEPA technical experts concerning the sediment bioassay study was taking place. This incongruity raises the question of how much effort was really put into reviewing the benthic community study by USEPA.

II. The sand reference stations used for the benthic community study were not appropriate because:

a. **There was significant mortality associated with the sediments from these stations in the sediment toxicity tests; and**
b. **Sediments from SQT10 and SQT12 exhibited, "a strong odor of decaying organic matter" and "elevated levels of ammonia" while in the sediment toxicity laboratory.**

These comments imply that these sand reference stations, and by extension SQT 13 and SQT14, additional sand reference stations which also had significant mortality in the bioassay testing, were somehow inappropriate as reference stations in the benthic community investigation because of what happened in the bioassay laboratory. However, examination of the benthic community data from these stations indicates that in spite of the observations made in the bioassay laboratory, i.e. odors in bioassay sediment, and the results of the bioassay testing, i.e. significant mortality, the benthic community structure at these reference stations was not significantly different from other reference or Site stations.

NSPW therefore questions the logic underlying USEPA's conclusion that the performance of reference stations in the bioassay should negate their use as reference stations in the benthic community investigation. In fact, USEPA's comments imply there was actually some ubiquitous variable that affected benthic community structure, even at sand reference stations. If so, it would follow that any effects seen in Site stations during the bioassay could have resulted from the influence of this variable. USEPA dismissed this as possibility in their interpretation of the bioassay results, however. USEPA can't have it both ways, ignoring the reference station bioassay results in interpreting the bioassays but also saying that the benthic community study was flawed because of the reference station results in the bioassay.

Consistent with the scientific method, the RI/FS Work Plan specified the use of reference stations. Sand Reference Stations SQT10 and SQT12 were located miles apart in Chequamegon Bay and Sand Reference Stations SQT13 and SQT14 were located miles from SQT10 and SQT12 and were sampled three months after the other sand reference stations. This should allay any concerns that there was some unique stressor present in a specific part of Chequamegon Bay that would disqualify samples from that area from consideration as a benthic community reference station. Additionally there were no observations of an odor or observation of any other stressors in the field samples collected for the benthic community study.

Reference stations are used in benthic community studies to ensure that effects seen at Site stations are in fact, attributable to Site contaminants. If the results from reference stations are internally consistent, as was the case in this benthic community study (and the bioassay study), the appropriate (sound science) reaction should be to conclude there is some variable affecting reference stations **and** Site stations, **not** to arbitrarily reject the reference stations results. The fact that USEPA arbitrarily dismisses the entire benthic community study on the tenuous and unsupportable grounds that the reference stations were inappropriate is contradictory to USEPA guidance (cf USEPA 2000) and the scientific method.

III. USEPA claims that NSPW erred by placing a high weight of evidence on the results of the benthic community study because according to the USEPA's *Guidance Manual to Support the Assessment of Contaminated Sediments in Freshwater Ecosystems – Volume III – Interpretation of the Results of Sediment Quality Investigations (EPA-905-BO2-001C, December 2002)*, "the information on benthic

community structure can not be used alone to evaluate the cause of any impacts observed." [and] "The "high weight of evidence" that NSPW attempts to place on the benthic macroinvertebrate community study is not supported by USEPA guidance".

At no time did NSPW suggest that the benthic community results be "used alone" to evaluate the cause of any impacts observed. However, NSPW did indicate that the benthic community study should be accorded the highest weight of evidence. This is consistent with the USEPA guidance cited by the reviewer which also says,

> 'In addition, field-collected sediments are manipulated before [toxicity] testing, which may affect their integrity and toxicity. Similarly, certain sediment phases (e.g., organic extracts, elutriates) may be less relevant for evaluating the *in situ* effects of toxic substances in sediments *(Volume III, page 35)*" and, "sediment toxicity testing should be included as an integral element [not the only]of most sediment quality assessments *(Volume III, page 42)*" and, the results of benthic invertebrate assessments should be [emphasis added] considered in conjunction with the results of the companion measures of sediment chemistry, sediment toxicity, and bioaccumulation that are conducted at the assessment area *((Volume III, page 58), and,* First and foremost, the results of these [benthic community] assessments provide information that is directly relevant for evaluating benthic invertebrate community status (i.e., evaluating the *in situ* effects of contaminated sediments on the benthic community). In addition, procedures for conducting such assessments have been established..... *(Volume III page 48)*".

While Volume III (page 52ff) of this guidance acknowledges the inherent variability of benthic communities, it also specifically describes and provides references on how careful experimental design and use of statistics can provide a scientifically sound approach to meet data quality objectives in evaluating the structure of benthic communities.

In addition USEPA Ecological Risk Assessment and Risk Management Principles for Superfund Sites (OSWER Directive 9285.7-28 P) indicates that, "Superfund remedial actions generally should not be designed to protect organisms on an individual basis (the exception being designated protected status resources, such as listed or candidate threatened and endangered species or treaty-protected species that could be exposed to site releases), but to protect local populations and communities of biota." (USEPA 1999).

Thus, not only USEPA guidance but also USEPA directive supports NSPW's decision to place a high weight of evidence to the results of the benthic community study. Indeed, USEPA's effort to minimize the results of benthic studies as an important line of evidence flies in the face of scientific consensus which indicates the "litmus test" for establishing causality between the presence of contaminants and effects to ecological receptors is to be able to document that effects predicted by sediment chemistry data and sediment toxicity data are actually manifested in the resident biotic community (USEPA 2000; Grapentine 2002; Wennig et al. 2005; Chapman 2007).

Summary

In summary the conclusion that the benthic community should be accorded the highest weight of evidence in the BERA is supported by the following factors:

1) The study was conducted using an USEPA-approved experimental design and suite of statistical methods that was based upon a power analysis conducted with site benthic data. Both *a priori* and a *posteriori* analysis confirmed that there was adequate power of these statistical analyses to detect a change due to PAH.
2) USEPA erred in rejecting the results of the benthic community study based upon the mistaken conclusion that since the sand reference stations experienced significant mortality in the sediment toxicity tests they were inappropriate to use as reference stations for analysis of benthic community structure.
3) USEPA guidance, USEPA directive and consensus of the scientific community is that the results of community studies are critical for establishing causality between the presence of contaminants and effects to ecological receptors in sediment studies.

NSPW believes USEPA failed to conduct a proper review of the benthic community study as well as the data that supported it. Unlike the comprehensive review given the sediment toxicity study by USEPA, the technical review of the benthic community study was inadequate. The comments on the benthic community study reveal that the reviewer(s) lacked the requisite skills, including knowledge of benthic community ecology, experimental design and statistical analysis of ecological data required for a sound scientific review. NSPW fails to understand why, after committing the time and resources to developing and implementing a RI Work Plan for study of the benthic community, USEPA focused scant attention to understanding how the experimental design and statistical analysis addressed the variability about which they are so concerned.

NSPW believes USEPA's directive to modify the scientific record of the RI without detailed scientific justification is arbitrary and, at the very least, a departure from USEPA's "Sound Science" policy which envisions that USEPA should be held to the same scientific standards to which the USEPA holds Responsible Parties. Additionally, the absence of a thorough review is inconsistent given the extraordinary effort USEPA expended to review and critique the sediment toxicity study.

Sincerely,

Jerry Winslow
Principal Environmental Engineer

Cc: Jamie Dunn, WDNR Project Manager
Rae Ann Maday, Bad River Band of Lake Superior Chippewa
Melonee Montano, Red Cliff Band of Lake Superior Chippewa

References

Chapman, P.M. 2007. Do not disregard the benthos in sediment quality assessments! Marine Pollution Bulletin 54 (2007) 633–635.

Grapentine, L., Anderson, J., Boyd, D., Burton, G.A., DeBarros, C., Johnson, G., Marvin, C., Milani, D., Painter, S., Pascoe, T., Reynoldson, T., Richman, L., Solomon, K., Chapman, P.M., 2002. A decision-making framework for sediment assessment developed for the Great Lakes. Human and Ecological Risk Assessment 8, 1641–1655.

Wenning, R.J., Ingersoll, C.G., Batley, C., Moore, M. (Eds.), 2005. Use of Sediment Quality Guidelines (SQGs) and Related Tools for the Assessment of Contaminated Sediments. SETAC Press, Pensacola, FL, USA.

USEPA. 1999. Ecological Risk Assessment and Risk Management Principles for Superfund Sites (OSWER Directive 9285.7-28 P)

USEPA. 2000. Stressor Identification Guidance Document. EPA/822/B- 00-025. U.S. Environmental Protection Agency, Office of Water and Office of Research and Development, Washington, DC, USA.

USEPA. 2002. Guidance Manual to Support the Assessment of Contaminated Sediments in Freshwater Ecosystems – Volume III – Interpretation of the Results of Sediment Quality Investigations (EPA-905-BO2-001C.

URS. 2005. Remedial Investigation/Feasibility Study Work Plan. Ashland/NPLWaterfront Superfund Site.